UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

A MULTIMODE MULTIPRODUCT NETWORK ASSIGNMENT MODEL FOR STRATEGIC PLANNING OF FREIGHT FLOW

by

Jacques Guélat,
Michael Florian, and
Teodor Gabriel Crainic

Cette recherche a été subventionnée par des octrois du C.R.D.I. et du C.R.S.N.G.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Université de Montréal

Centre de recherche sur les transports - Publication #549

Département d'Informatique et de
Recherche opérationnelle - Publication #637

March 1987

Dépôt légal - 4e trimestre 1987

ABSTRACT

We present in this paper a normative model for simulating freight flows of multiple products on a multimodal network. The multimodal aspects of the transportation system considered are accounted for in the network representation chosen. The multiproduct aspects of the model are exploited in the solution procedure, which is a Gauss-Seidel - Linear Approximation Algorithm. An important component of the solution algorithm is the computation of shortest paths with intermodal transfer costs. Computational results obtained with this algorithm on a network that corresponds to the Brazil transportation network are presented.

Key words : multicommodity flows, Gauss-Seidel method, linear approximation algorithm, shortest path with turning penalties.

RÉSUMÉ

Nous présentons dans cet article un modèle normatif pour la prévision des flux des marchandises de plusieurs produits sur un réseau multimodal. Les aspects multimodaux du système de transport considéré sont reflétés dans la représentation du réseau. Les aspects multiproduits sont exploités dans la procédure de solution qui est un algorithme du type Gauss-Seidel-Approximation linéaire. Une composante importante de l'algorithme de solution est le calcul des chemins plus courts avec des pénalités de transfert intermodal. Des résultats obtenus avec cet algorithme sur un réseau qui correspond au réseau national du Brésil sont présentés.

Mots-clés : flots à multicommodités, méthode de Gauss-Seidel, algorithme d'approximation linéaire, plus courts chemins avec pénalités de transfert.

1.0 Introduction

The prediction of multicommodity freight flows over a multimodal network has attracted much interest in recent years. In contrast to urban transportation, where the prediction of passenger flows over multimodal networks has been studied extensively and many of the research results have been transferred to practice (see, for instance, Florian, 1984), the study of freight flows at the national or regional level, perhaps due to the inherent difficulties and complexities of such problems, received less attention.

The class of models that was well studied in the past for prediction of interregional flows is the spatial price equilibrium model and its variants. The model, stated initially by Samuelson (1952) and extended by Takayama and Judge (1964, 1970) then by Florian and Los (1982), Friesz, Tobin, Smith, and Harker (1983) has been used extensively for analyzing interregional commodity flows. This class of models determines simultaneously the flows between producing and consuming regions as well as the selling and buying prices. The transportation network is usually modeled in a simplistic way and these models rely to a large extent on the supply and demand functions of the producers and consumers respectively. The calibration of these functions is essential to the application of these models and the transportation costs are unit costs or may be functions of the flow on the network. There have been so far few multicommodity applications of this class of models, with the majority of applications having been carried out in the agricultural and energy sectors in an international or interregional setting. It is not this class of models which is the main topic of this paper.

The class of models which we consider are network models which enable the prediction of multicommodity flows over a multimode network, where the physical network is modeled at a level of detail appropriate for a nation or a large region and represents the physical facilities with relatively little abstraction. The demand for the transportation services

is exogenous and may originate from an input-output model, if one is available, or from other sources, such as observed demand or scaling of observed past demand. The choice of mode, or subsets of modes used is exogenous and intermodal shipments are permitted. In this sense, these models may be integrated with econometric demand models as well. The emphasis is on the network representation and the proper representation of congestion effects in a static model aimed to serve for comparative static analyses or for discrete time multiperiod analyses.

The Harvard Model (Kresge and Roberts, 1971), which is probably the first published predictive freight network model of the type that interests us, resorted to a fairly simple "direct link" representation of the physical network and congestion effect were not considered. Later, the Multi-State Transportation Corridor Model (McGinnis et al., 1981; Jones and Sharp, 1979; and Sharp, 1979) went a step further in representing an explicit multimodal network, but without any consideration of congestion. The Transportation Network Model (Bronzini, 1980) does not consider congestion effects either. The first model that considers congestion phenomena in this field, is the Freight Network Equilibrium Model (Friesz, Gottfried and Morlok, 1986). This is a sequential model which uses two network representations: an aggregate network that is perceived by the users, which serves to determine the carriers chosen by the shippers and then more detailed separate networks for each carrier, where commodities are transported at least total cost.

In an effort to combine the variable demand modeling approach of spatial equilibrium models with detailed description of the shippers and carriers, in recent articles, Harker and Friesz (1986a, 1986b) present a theoretical approach and mathematical formulations that are yet to be tested in a particular application.

The class of models that we present in this paper do not consider shippers and carriers as distinct actors in the decisions made in shipping freight. The simplifying assumption is made that the goods are shipped at least cost, even though the shipment of freight is governed by a variety

of micro-circumstances that prevent the achievement of total cost minimization. The justification for this assumption is that, at a strategic level, predicting freight flows based on this assumption, when reliable modal demand is available, is satisfactory for the purpose of scenario comparisons when the investments considered are of large magnitude.

The multimodal aspects of a national transportation system are accounted for in the network representation chosen. A link of the multimodal network is defined by its origin and destination nodes and a single mode. Parallel links are allowed between two adjacent nodes, one for each mode available to transport goods between them. The intermodal transfers at a node of the network are modeled as link to link permitted movements. Appropriate cost functions may be associated with links and intermodal transfers.

The multiproduct aspects of a national transportation system are accounted for in the formulation of the predictive model and is taken advantage of in the solution procedure. The algorithm developed for this problem exploits the natural decomposition by product and results in a Gauss-Seidel like procedure.

The multimode-multiproduct model is formulated in the most general way, permitting in principle nonconvex and assymmetric cost functions. Nevertheless, certain simplifying assumptions made on the structure of the cost functions may simplify the problem and permit the solution of large size problems in reasonable computational times.

The paper, which is mainly based on Guélat's (1986b) work, is organized in the following way. First, the network representation chosen to integrate modes and transfers is described in detail; then the multimode multiproduct assignment model is formulated and the solution algorithm adopted is presented; next, two salient aspects of the model are analyzed in detail, namely the form of the average and marginal cost functions and the computation of shortest paths with intermodal transfer costs on the network representation chosen; finally, some computational results obtained with this algorithm are presented.

2.0 The Network Representation

The physical network infrastructure represented by the network model

chosen supports the transportation of several products on several modes. A product is any commodity (collection of similar products), goods or passengers, that generates a link flow specifically associated with it. A mode is a means of transportation that has its own characteristics, such as vehicle type and capacity as well as a specific cost function.

The base network is the network that consists of the nodes, links and modes that represent all the physical movements possible on the available infrastructure. The model that we have chosen defines a link as a triplet (i,j,m) where i is the origin node, $i \in N$, where N is the set of nodes of the network, j is the destination node, $j \in N$ and m is the mode allowed on the arc, $m \in M$, where M is the set of modes available on the network. Parallel links are used to represent the situation where more than one mode is available for tranporting goods between two adjacent nodes. In order to motivate our network modeling choice, consider the following simple example, which is shown in Figure 1. It consists of three modes: road transport, diesel rail and electrified rail.

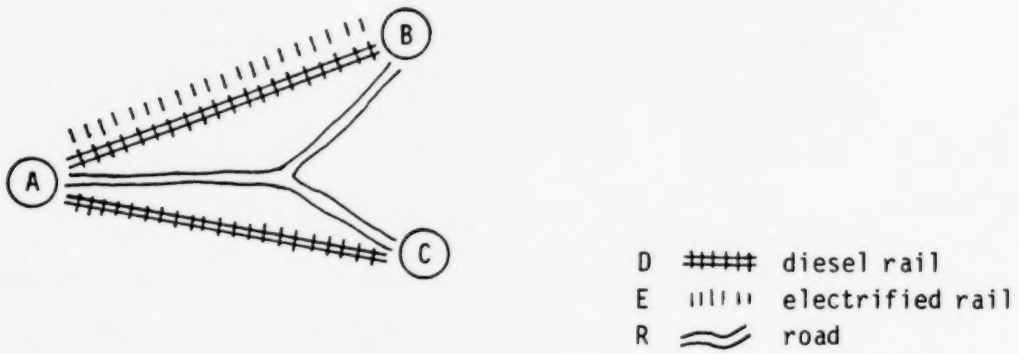


Figure 1: Physical Network

Cities A and B are served by all modes; A and C are served by diesel rail and road, whereas B and C are served only by road. In order to simplify the presentation, no intermodal transfers are permitted.

The most compact representation of this network is to connect all cities by directed links and to allocate the modes as link attributes.

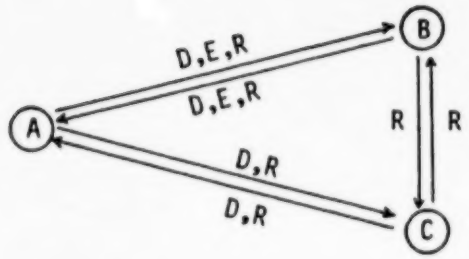
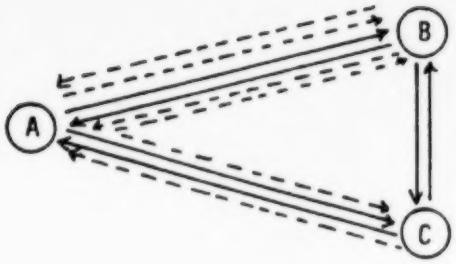


Figure 2: Compact Representation

Thus, link (A,B) has all the modes permitted, link (A,C) has the diesel and road modes, while link (B,C) has only the road mode. This network representation has some major disadvantages for the model that we seek. If a single flow is associated with a link, it must be the total flow on the link and the flows specific to each mode are not kept explicitly. If multiple flows, one for each mode, are associated with each link, then these may vary in number from link to link, depending on the number of modes associated with a given link. In addition, the physical difference of the infrastructure modeled is not explicit in such a representation and the specification of cost and delay functions for each mode poses the same problem as that mentioned above for the flows for each mode. In order to overcome these drawbacks, it is necessary to choose a network representation which permits easily the identification of flow of goods by mode, as well as cost and delay functions by mode. This is equivalent to keeping explicitly "copies" of the network, one for each mode defined in a particular application. This type of network representation becomes prohibitive if large networks with many modes need to be considered. An elegant way of achieving the aim of efficient representation of the multimodal network is to permit parallel links between each node, one for each mode permitted. In this way, the network model resembles the physical network, since the rail and road infrastructures are physically different. Also, if on a physical link, such as the rail facility, there are two different types of services, such as the diesel and electric train services of the example described in Figure 1, a separate link may be attributed to each service, since they would have different cost and delay functions. For this example, the network with parallel links is described in Figure 3.

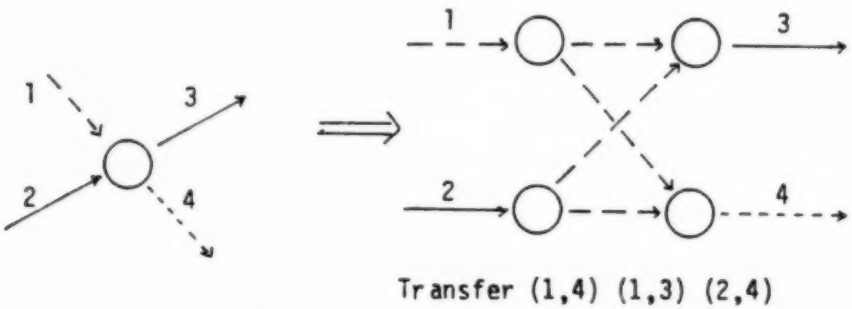


——— road
– – – – rail diesel
------- rail electric

Figure 3: Network Representation Chosen

In this representation, the mode is an integral part of the network and is not merely an attribute of the link. The parallel links introduced in the network, as may be seen in Figure 3, may lead to difficulties in implementing network algorithms unless parallel link between adjacent nodes that represent different modes are structurally different. This is accomplished by defining a link as a triplet (origin, destination, mode).

Once the network representation chosen, it is necessary, in order to model intermodal shipments, to permit and associate the appropriate costs and delays for mode to mode transfers at certain nodes of the network. This may be accomplished by expanding (or exploding) a node where transfers occur by adding as many nodes as there are arcs entering and exiting the node and by adding transfer links between these nodes to form a bipartite graph. An example is given in Figure 4.



——— road
– – – – rail diesel
------- rail electric

Figure 4: Explosion of a node

The explicit explosion of nodes would increase the number of links and arcs of the network; also, if carried out explicitly, the new nodes do not represent the physical network and the transfer arcs represent a mode to mode transfer so they are not unimodal as the other arcs of the network.

The representation chosen for transfers does not require the explicit modification of the network. Transfers are represented implicitly by a pair of arcs, one reaching the node and the other leaving the node. Transfer arcs are not added to the base network; rather the transfers permitted at each node may be displayed or listed by referring only to the pairs of arcs defining the transfers at a given node.

3.0 The Multimode Multiproduct Model

In the context of strategic planning of freight flows on a national or regional scale, the most efficient use of the transportation infrastructure is to transport the freight at least total (generalized) cost. Even though it is reasonable to assume, that even in countries where a central authority controls and regulates the shipment of goods, a variety of micro-circumstances in fact prevent the precise achievement of the goal of shipping at least cost, the model that we formulate is based on the objective of minimizing total costs. The notion of cost is central to our model and we interpret cost in the most general way, in the sense that it may have different components, such as monetary cost, delay, energy consumption, etc.

The network that we consider consists of a set of nodes N, a set of arcs A, A $\subset$ NxN, a set of modes M and a set of transfers T, T $\subset$ AxA. We denote their cardinality, n_N, n_A, n_M and n_T respectively. With each arc a, a$\in$A, we associate a cost function $s_a(\cdot)$ which depends on the volume of goods on the arc, or possibly, on the volume of goods on other arcs of the network. Similarly, a cost function $s_t(\cdot)$ is associated with each transfer t$\in$T.

The products transported over the multimodal network are denoted by index p, p$\in$P, where P is the set of all products considered, which is of cardinality n_p. Each product is shipped from origins o, o$\in$O $\subseteq$ N, and destinations d, d$\in$D $\subseteq$ N of the network. The demand for each product and for all origin/destination (O/D) pairs is specified by a set of O/D matrices for the corresponding subsets of modes. It is assumed that the mode choice is

determined exogenously. Let $g^{m(p)}$ be the demand matrix associated with product $p\varepsilon P$, where $m(p)$ is a subset of modes that belong to $M(p)$, the set of all subsets of modes that are used to transport product p.

The flows of product p on the multimodal network is denoted by v^p and consists of the induced flows of this product on links and transfers.

$$v^p = \begin{pmatrix} (v_a^p),\ a\varepsilon A \\ (v_t^p),\ t\varepsilon T \end{pmatrix}$$

The flow of all the products on the multimodal network is denoted by $v=(v^p)$, $p\varepsilon P$ and is a vector of dimension $n_p(n_A+n_T)$.

The average cost functions $s_a^p(v)$, on links, and $s_t^p(v)$, on transfers, correspond to a given flow vector v. The average cost functions for product p are denoted, similar to the notation used for the flow v^p, s^p, $p\varepsilon P$, where

$$s^p = \begin{pmatrix} (s_a^p),\ a\varepsilon A \\ (s_t^p),\ t\varepsilon T \end{pmatrix}$$

and $s=(s^p)$, $p\varepsilon P$ is the vector of average cost functions of dimension $n_p(n_A+n_T)$.

The total cost of the flow on arc a, $a\varepsilon A$, for the product p, $p\varepsilon P$ is the product $s_a^p(v)v_a^p$; the total cost of the flow on transfer t, $t\varepsilon T$ is $s_t^p(v)v_t^p$. The total cost of the flows of all products over the multimodal network is the function F that we seek to minimize

$$F = \sum_{p\varepsilon P}\left(\sum_{a\varepsilon A} s_a^p(v)v_a^p + \sum_{t\varepsilon T} s_t^p(v)v_t^p\right) \quad (=s(v)^T v) \qquad [1]$$

over the set of flows which satisfy the conservation of flow and nonnegativity constraints. In order to write these constraints for the multiproduct multimode network defined above, the following notation is used. Let $K_{od}^{m(p)}$ denote the set of paths that lead from origin o, $o\varepsilon O$ to destination d, $d\varepsilon D$, by using only modes of $m(p)\varepsilon M(p)$, $p\varepsilon P$. The conservation of flow equations are then

$$\sum_{k\varepsilon K_{od}^{m(p)}} h_k = g_{od}^{m(p)}, \qquad o\varepsilon O,\ d\varepsilon D,\ m(p)\varepsilon M(p),\ p\varepsilon P \qquad [2]$$

where h_k is the flow on path k. The nonnegativity constraints are

$$h_k \geq 0, \qquad k \in K_{od}^{m(p)},\ o \in O,\ d \in D,\ m(p) \in M(p),\ p \in P \qquad [3]$$

Let Ω be the set of flows v that satisfy [2] and [3]. Since the conservation of flow equations are stated in the space of path flows, for notational convenience, the specification of Ω requires the relation between arc flows and path flows, which is

$$v_a^p = \sum_{k \in K^p} \delta_{ak} h_k, \qquad a \in A,\ p \in P \qquad [4]$$

where $K^p = \bigcup_{m(p) \in M(p)} \bigcup_{o \in O} \bigcup_{d \in D} K_{od}^{m(p)}$ is the set of all paths that may be used by product p, and

$$\delta_{ak} = \begin{cases} 1 \text{ if } a \in k \\ 0 \text{ otherwise} \end{cases}$$

is the indicator function which identifies the arcs of a particular path. Similarly, the flows on tranfers are

$$v_t^p = \sum_{k \in K^p} \delta_{tk} h_k, \qquad t \in T,\ p \in P \qquad [5]$$

where $\delta_{tk} = \begin{cases} 1 \text{ if } t \in k \\ 0 \text{ otherwise} \end{cases}$.

The transfer t belongs to the path k if the two arcs that define the transfer belong to it.

In conclusion, the system optimal multiproduct, multimode assignment model consists of minimizing [1] subject to [2]-[3] with the definitional constraints [4]-[5].

The model is sufficiently general to be adapted for different ways of specifying the demand. Even though typical applications of the model are likely to be carried out after an "a priori" mode choice calculation, which would allocate the demand for a product g^p to a set of mode subsets, it is equally possible to permit the demand for a product to be transported over all the allowed modes, that is m(p) is the set of all modes of the network and M(p) has a single component, which in this case is m(p). Also, the model is flexible in the specification of intermodal movements. The mode to

mode transfers may be restricted to occur only at specific nodes of the network, and only between specific modes.

We turn our attention next to develop a solution algorithm for this model.

4.0 A Solution Algorithm

The problem that we consider may be stated in a compact form as

$$\text{Min } F(v), \text{ subject to } v \in \Omega \qquad [P]$$

Since constraints [2]-[5] are linear, Ω is a polytope which, if the multimodal network is strongly connected, is nonempty. Ω is compact as well, if the cost functions that we consider are strictly positive, hence the paths considered do not contain cycles.

It is assumed that $F(v)$ is a convex function, once differentiable on an open space that contains Ω. These assumptions are satisfied, for example, by average arc (transfer) cost functions of the type

$$s_a^p = \sum_i \alpha_i (\beta_i + v_a^p)^{z(i)}, \qquad \alpha_i > 0,\ \beta_i > 0,\ 0 \leq z(i) < \infty \quad .$$

As demonstrated by this example, the average cost functions need not be convex ($0 < z(i) < 1$) for $F(v)$ to be convex.

The size of the problems that are likely to occur in specific applications ($n_p(n_A+n_T) \sim 100{,}000$-$150{,}000$) mitigates the use of second order algorithms that require the use of the Hessian of the objective function or its approximation in order to solve [P]. The most widely used first order method is the linear approximation method (Frank and Wolfe, 1956). This method, which is widely used for solving large scale network equilibrium problem (see Florian, 1986) has a sublinear rate of convergence in the vicinity of the optimal solution, however it has proved to be robust in applications which do not require a very precise solution. Its straightforward adaptation for the model that we seek to solve poses a problem of dimension: the vector v is of dimension $n_p(n_A+n_T)$ and so is the dimension of the feasible direction of descent generated by the linear approximation method. This may render the straightforward application of the linear approximation method impracticable on the most widely available computers.

The structure of the model suggests a natural decomposition by product. Ω is the direct product $\prod_p \Omega^p$, where Ω^p is the set of feasible flows of product p on the subnetworks $m(p) \in M(p)$. The descent direction of the linear approximation method is found by minimizing the linear approximation of the objective function, which in this case would be F^p ($F = \sum_{p=1}^{n_P} F^p$); thus one would have to solve n_p subproblems in a cycle of the algorithm. The algorithm that we develop in this section is based on this observation, and may be characterized as a Gauss-Seidel - Linear Approximation Approach. For each product p, a descent direction is computed for v^p which is of dimension (n_A+n_T). When considering a particular product, all the flow variables pertaining to the other products are kept at their previous value.

This approach resembles the block Gauss-Seidel method, except that the subproblem for each block is not solved exactly; rather, for each product p, a single iteration of the linear approximation algorithm is carried out. The statement of the algorithm is as follows:

Algorithm A

Step 0 - Initialization

Determine v (initial feasible solution)

Step 1 - Major cycle

$v_{old} = v$; for each $p \in P$; carry out 1p.

1p. Minor cycle

Compute $c^p(v)$	(marginal costs)
Find w^p	(extreme point found by shortest path computations)
$d^p = w^p - v^p$	(descent direction)
Compute λ_p	(optimal step size)
$v^p = v^p + \lambda_p d^p$	(update flow for product p).

Step 2 - Stopping criterion

If $(v \neq v_{old})$, return to step 1.

Since the algorithm is stated in a rather compact form, some comments on the details of each step are in order. The minor cycle 1p. implements one iteration of the linear approximation method in the subspace of flows related to product p. A descent direction is found by minimizing the linear approximation of F^p on Ω^p, which is

$$\text{Min} \quad \nabla F^p(v)y, \text{ subject to } y\epsilon\Omega^p \qquad (LP_p) \ .$$

Since $F(v)=s(v)^T v$ it follows that $\nabla F(v) = s(v)^T + v^T s'(v)$. Let the marginal cost be defined as $c(v) \doteq s(v) + s'(v)^T v$. The linear programming subproblem is

$$\text{Min} \quad c^p(v)^T y, \text{ subject to } y\epsilon\Omega^p \ .$$

The solution of this problem, as is well known, is obtained by assigning the demand $g^{m(p)}$ on the shortest paths corresponding to the modes that belong to m(p) computed with arc costs $c^p(v)$.

The optimal step size λ_p is obtained by a one-dimensional minimization of F in the direction $(0,\ldots,d^{p^T},\ldots,0)^T$ unless $c^p(v)^T d^p=0$ in which case $\lambda_p=0$.

An initial (extremal) solution may be otained by carrying out a major cycle with initial marginal costs corresponding to v=0, that is c(o) and by setting $\lambda_p=1$ in each minor cycle 1p.

The algorithm is stopped when the flows v do not change after a major cycle was carried out. The current solution is optimal, as will be demonstrated in the following section. In practice, the algorithm may be stopped after carrying out a predetermined number of major cycles or after a maximal execution time. In addition, the supposed convexity of the objective function F may be used to construct a stopping criterion based on the "relative gap" RG

$$RG = 100 * \frac{F(v)-BLB(v)}{F(v)}$$

where BLB(v) represents the best lower bound obtained during the previous major cycles carried out. This lower bound is computed by solving the linear programming problem

Min $\nabla F(v)^T y$, subject to $y \in \Omega$.

A lower bound on the optimal solution $F(v^*)$ is given by $F(v)+\nabla F(v)(y^*-v)$, where y^* is the optimal solution of the above problem. This lower bound may not be computed by using the shortest paths detained at each minor cycle 1p.; it requires the solution of as many shortest paths problems as there are O/D matrices $g^{m(p)}$. As it is rather "expensive" to compute, its computation would depend on the power of the computer available for solving the problem.

5.0 Convergence Analysis of the Algorithm

In order to prove the global convergence of the algorithm presented in the previous section, we show that F is a descent function relative to algorithm A and that the algorithmic map, which we denote by A as well, is closed on Ω.

Lemma 1: Let y be the solution obtained from v after carrying out a minor cycle $1\bar{p}$. $F(y)<F(v)$.

Proof:
$$y^p = \begin{cases} v^p + \lambda_p d^p & \text{if } p=\bar{p} \\ v^p & \text{otherwise.} \end{cases}$$

Then $y=v+\lambda_p D$, where $D=(D^p)_{p\in P}$ and $D^p = \begin{cases} d^p=w^p-v^p, & \text{if } p=\bar{p} \\ 0 \text{ otherwise} . \end{cases}$

Let $g(\lambda)=F(v+\lambda D)$; by the assumptions made on F, g is continuously differentiable. Thus

$$g'(\lambda) = \nabla F(v+\lambda D)D = c(v+\lambda D)^T D$$

$$= \sum_{p\in P} c^p(v+\lambda D)^T D^p = c^p(v+\lambda D)^T d^p$$

and $g'(o) = c^p(v)^T d^p$

- if $g'(0)=0$, then $\lambda_p=0$ in which case $y=v$ and $F(y)=F(v)$.
- if $g'(0)\neq 0$, then $g'(0)<0$ by the definition of w^p; thus D is a feasible descent direction for F at v. By the definition of λ_p, it follows that $F(y)<F(v)$.∎

Now let $S \subseteq \Omega$ be set of flows that satisfy the Kuhn Tucker conditions of problem (P), or equivalently, the set of optimal solutions of P. If F is strictly convex S is a singleton. The set S may be characterized by the variational form

$$S = \{v \in \Omega \mid \nabla F(v)(w-v) \geq 0, \quad \forall w \in \Omega\} .$$

Let A: $\Omega \to \Omega$ be the (one to many) algorithmic map.

Lemma 2: $v \in S$ if and only if $v \in A(v)$.

Proof: First we prove that the condition is necessary.

$$\nabla F(v)(w-v) \geq 0 \quad , \quad \forall w \in \Omega \qquad \text{is equivalent to}$$

$$\sum_{p \in P} c^p(v)^T (w^p - v^p) \geq 0, \quad \forall w \in \Omega .$$

Consider $\bar{p} \in P$ and $y = (y^p)_{p \in P} \in \Omega$ such that $y^p = v^v$ if $p \neq \bar{p}$. It follows that

$$c^{\bar{p}}(v)^T (y^{\bar{p}} - v^{\bar{p}}) \geq 0, \quad \forall y^{\bar{p}} \in \Omega^{\bar{p}} .$$

In particular, if $w^{\bar{p}}$ is the extremal flow generated in the linear subproblem of the minor cycle $1\bar{p}$, it follows that $c^{\bar{p}}(v)^T (w^{\bar{p}} - v^{\bar{p}}) = 0$. Hence v will not be modified in any minor cycle and $v \in A(v)$.

The sufficiency of the condition is demonstrated by contraposition. Let $v \in \Omega - S$. There exists then a $w \in \Omega$ such that

$$\nabla F(v)(w-v) < 0 \quad \text{or} \quad \sum_{p \in P} c^p(v)^T (w^p - v^p) < 0 .$$

It is impossible for $c^p(v)^T d^p = 0$ for all $p \in P$; but, in the proof of Lemma 1 it was shown that if $c^p(v)^T d^p < 0$ at a minor cycle the objective function F decreases strictly and as F cannot increase at a minor cycle, it follows that $F(y) < F(v)$ for all $y \in A(v)$. In particular v does not belong to A(v).∎

Corollary 1: F is a function of descent relative to A.

Proof: Let $y \in A(v)$. By applying Lemma 1 n_p times, one for each subproblem p, it follows that $F(y) \leq F(v)$. By Lemma 2, $F(y) < F(v)$ if $v \in \Omega - S$. ∎

The second part of the convergence proof consists of the demonstration that the algorithm map A is closed on Ω. In order to do so, let $P=(p_1,\ldots,p_{n_p})$ and A is decomposed accordingly:

$$A = A^{p_{n_p}} \circ \ldots\ldots\ldots \circ A^{p_1} : \Omega \to \Omega$$

where $A^p = M \circ D^p : \Omega \to \Omega$

A^p is the algorithmic map that corresponds to a minor cycle, D^p represents the search for a descent direction and M the one-dimensional minimization. More formally, let

$$D^p : \Omega \to \Omega \times U$$
$$v \to D^p(v)$$

where $D^p(v) = \{(v,d) \mid d^T = (0,\ldots,d^{p^T},\ldots 0),\ d^p = w^p - v^p$, where

$$w^p \in \underset{y \in \Omega^p}{\text{Arg Min}}\ \nabla F^p(v)y\},$$

and $U = \Omega - \Omega = \{y \mid y = y_1 - y_2;\ y_1, y_2 \in \Omega\}$

$$M : \Omega \times U \to \Omega$$
$$(v,d) \to M(v,d)$$

where $M(v,d) = \{y \mid y = \begin{cases} v & \text{if } \nabla F(v)d = 0 \\ v+\lambda^* d & \text{otherwise, with } \lambda^* \in \underset{0<\lambda<1}{\text{Arg Min}}\ F(v+\lambda d)\}. \end{cases}$

Lemma 3: D^p is a closed mapping on Ω.

Proof: Let $\{v\}_k$ be a sequence of Ω that converges to v. Let, on the other hand, $\{(v^k, d^k)\}_k$ be a sequence of $\Omega \times U$ such that $(v^k, d^k) \in D^p(v^k)$ and which converges to (v,d). By the definition of $D^p(v^k)$ the direction d^k may be expressed as $y^k - v^k$ where $y^k = v^k$ except for

the p'th component where $(y^k)^p = (w^k)^p$. Thus we have that

$$y^k = d^k + v^k, \quad y^k \in \Omega .$$

The right term of the equation tends to $d+v \doteq y$; $y=v$ except for the p'th component. In order to show that (v,d) belongs to $D^p(v)$ it suffices to point out that y^p is the optimal solution of the linear subproblem

$$\text{Min} \quad \nabla F^p(v)z, \quad \text{subject to } z \in \Omega^p .$$

Then $\nabla F^p(v^k)(y^k)^p \leq \nabla F^p(v^k)z^p, \quad \forall z^p \in \Omega^p .$

By invoking the continuity of ∇F it follows that

Then $\nabla F^p(v)y^p \leq \nabla F^p(v)z^p, \quad \forall z^p \in \Omega^p .$ ∎

Lemma 4: The map M is closed on $D^p(v)$, $\forall v \in \Omega$.

Proof: Let $(v,d) \in D^p(v)$ and let $\{(v^k, d^k)\}_k$ be a sequence of $\Omega \times U$ which converges to (v,d). On the other hand, let $\{y^k\}_k$ be a sequence of Ω such that $y^k \in M(v^k, d^k)$, which converges to y. Consider first the case where $d \neq 0$. There exists then a K such that $d^k \neq 0$ for all $k > K$. In the following we consider that $k > K$

$$y^k = v^k + \lambda^{*k} d^k ,$$

$$\text{hence } \lambda^{*k} = \frac{\| y^k - v^k \|}{\| d^k \|} .$$

The term on the right hand side tends to

$$\frac{\| y-v \|}{\| d \|} \doteq \lambda^* (\in [0,1])$$

and hence $y = v + \lambda^* d$.

It remains to show that the λ^* above solves the one-dimensional minimization problem. Since $F(y^k) \leq F(v^k + \lambda d^k)$ for all $\lambda \in [0,1]$, by using the continuity of F and by taking the limit for fixed λ, one obtains:

$$F(y) = F(v + \lambda^{*}d) < F(v + \lambda d), \quad \lambda \in [0,1].$$

We consider now the case d=0.

$$y^k - v^k = \lambda^{*k} d^k .$$

As $\lambda^{*k} \in [0,1]$ for all k and as d^k tends to zero, the term on the right hand side tends to zero as well. By taking the limit one obtains that y=v.∎

Corollary 2: A^p is closed on Ω.

Proof: $A^p : \Omega \to \Omega x U \to \Omega$

$\Omega x U$ is compact since Ω is compact. By Lemmas 3 and 4 all the hypotheses of Lemma 4.2 of Zangwill (1969) are satisfied.∎

Corollary 3: A is closed on Ω.

Proof: By applying corollary 2 to $A=A^{p_n} p_o \ldots\ldots o A^1$ and Lemma 4.2 of Zangwill (1969) the result follows ∎

The global convergence of the algorithm A follows from convergence Theorem A of Zangwill (1969) and Corollaries 1 and 3 above.

We turn our attention next to the average and marginal cost functions.

6.0 The Cost Functions

In the minimization problem [P], defined in section 4 of this paper, the average cost functions $s_a(v)$, $a \in A$ and $s_t(v)$, $t \in T$ may in principle depend on any or all components of the vector v. The Jacobian matrix, which appears in the definition of the marginal costs, may therefore be a full matrix, which usually implies lengthy evaluations of these costs. We consider next several restrictive and simplifying hypotheses in the structure of the average cost functions, which preserve nevertheless sufficient generality to model properly nonseparable average arc cost functions.

In its most general form, the marginal cost for transporting product $\bar{p}$ on arc $\bar{a}$ is given by the expression

$$c^{\bar{p}}_{\bar{a}} = s^{\bar{p}}_{\bar{a}} + \sum_{p\epsilon P}\left(\sum_{a\epsilon A} \frac{\partial s^{p}_{a}}{\partial v^{\bar{p}}_{\bar{a}}} v^{p}_{a} + \sum_{t\epsilon T} \frac{\partial s^{p}_{t}}{\partial v^{\bar{p}}_{\bar{a}}} v^{p}_{t}\right) \qquad [6]$$

and the marginal cost for transporting product $\bar{p}$ on transfer arc $\bar{t}$ is

$$c^{\bar{p}}_{\bar{t}} = s^{\bar{p}}_{\bar{t}} + \sum_{p\epsilon P}\left(\sum_{a\epsilon A} \frac{\partial s^{p}_{a}}{\partial v^{\bar{p}}_{\bar{t}}} v^{p}_{a} + \sum_{t\epsilon T} \frac{\partial s^{p}_{t}}{\partial v^{\bar{p}}_{\bar{t}}} v^{p}_{t}\right) \qquad [7]$$

The first simplifying assumptions that we make concern the dependence of the cost function on an arc on the flows on the transfers and vice versa.

Assumption 1: $s^{p}_{t}(v)$, $p\epsilon P$, $t\epsilon T$ do not depend on the arc flows v^{p}_{a}, $p\epsilon P$, $a\epsilon A$. With this assumption equation [6] simplifies to

$$c^{\bar{p}}_{\bar{a}} = s^{\bar{p}}_{\bar{a}} + \sum_{p\epsilon P}\sum_{a\epsilon A} \frac{\partial s^{p}_{a}}{\partial v^{\bar{p}}_{\bar{a}}} \cdot v^{p}_{a} \qquad [8]$$

A similar simplification of equation [7] results from:

Assumption 2: $s^{p}_{a}(v)$, $p\epsilon P$, $a\epsilon A$ do not depend on the transfer flows v^{p}_{t}, $p\epsilon A$, $t\epsilon T$, which leads to the expression

$$c^{\bar{p}}_{\bar{t}} = s^{\bar{p}}_{\bar{t}} + \sum_{p\epsilon P}\sum_{t\epsilon T} \frac{\partial s^{p}_{t}}{\partial v^{\bar{p}}_{\bar{t}}} \cdot v^{p}_{t} \qquad [9]$$

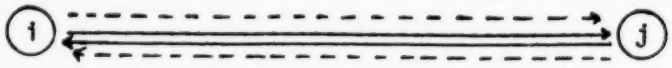
Since transfers represent mode to mode movements it is reasonable to postulate

Assumption 3: s^{p}_{t}, $p\epsilon P$, $t\epsilon T$ do not depend on the flows of the other transfers $\bar{t}\epsilon T$, $\bar{t}\neq t$.

[9] simplifies further with this assumption to

$$c^{\bar{p}}_{\bar{t}} = s^{\bar{p}}_{\bar{t}} + \sum_{p \in P} \frac{\partial s^{p}_{\bar{t}}}{\partial v^{\bar{p}}_{\bar{t}}} \cdot v^{p}_{t} \qquad [10]$$

Unfortunately, a similar assumption cannot be made for cost functions of the arcs. It is highly likely that two distinct modes (such as rail diesel, rail electric on a single track facility) share the same physical infrastructure between two nodes of the network. It is reasonable nevertheless to assume that the interactions are limited to links which are parallel, in both orientations, between two adjacent nodes. See Figure 5.



mode m mode m'

Figure 5: Parallel Links in Both Orientations

We introduce some additional notation. The arc $a=(i,j,m)$ connects node i to node j by mode m. Let M_a be the subset of modes which is considered in the computation of the marginal cost of arc a.

Let $A_a \doteq \{\tilde{a} \in A \mid \tilde{a} = (i,j,m') \text{ or } \tilde{a} = (j,i,m'),\ m' \in M_a\}$

be the set of arcs that must be considered in the computation of the marginal cost of arc a; thus $s^{p}_{a}(v) = s^{p}_{a}(v^{\tilde{p}}_{\tilde{a}})$, $\tilde{a} \in A_a$, $\tilde{p} \in P$. If arc $\tilde{a} \in A_a$ then $\frac{\partial s^{p}_{a}}{\partial v^{\tilde{p}}_{\tilde{a}}} = 0$. This permits the simplification of equation [8] to

$$c^{\bar{p}}_{\bar{a}} = s^{\bar{p}}_{\bar{a}} + \sum_{p \in P} \sum_{a \in H_{\bar{a}}} \frac{\partial s^{p}_{a}}{\partial v^{\bar{p}}_{\bar{a}}} \cdot v^{p}_{a} \qquad [11]$$

where $H_{\bar{a}} = \{a \in A \mid \bar{a} \in A_a\}$. The set $H_{\bar{a}}$ can be obtained in general by scanning the arcs parallel to $\bar{a}$ in both orientations, which may be time consuming particularly if $H_{\bar{a}} = \{\bar{a}\}$ for the majority of the arcs. Therefore we make the additional

Assumption 4: $H_a = A_a$ for all $a \in A$.

The marginal costs of the arcs may be writen then as

$$c_{\bar{a}}^{\bar{p}} = s_{\bar{a}}^{\bar{p}} + \sum_{p \in P} \sum_{a \in A_{\bar{a}}} \frac{\partial s_a^p}{\partial v_{\bar{a}}^{\bar{p}}} \cdot v_a^p \qquad [12]$$

In the code which implements the algorithm of the preceeding section in STAN, the partial derivatives required in the computation of the marginal costs are carried out by a rather precise numerical approximation procedure. The analytic form of the cost functions may be rather complex, in particular for the rail mode (see Crainic et al.), and as a consequence, so are the analytic form of the partial derivatives and cost functions. Another consideration that motivates the numerical approximation procedure is that it obviates the specification of the partial derivatives by the user of the code, be they simple or complex.

The method that we chose to implement in order to compute the derivative of a function at a given value, is a polynomial approximation based on the least squares method, which appears to be superior to numerical interpolation (Kellison (1975), pp. 150-151). We have chosen to use the orthogonal polynomial of Gram, which was adapted for our purposes (Guélat, 1986a) starting from a code written by Casaletto and Rice (1978). The interested reader is referred to Guélat for the details of the numerical approximation procedure for computing partial derivatives.

We turn our attention to the shortest path algorithm used to compute least marginal cost paths, which has a central role in the solution algorithm of the multimode multiproduct model.

7.0 The Shortest Path Algorithm

The multimodal network that we use for the formulation of the multi-product assignment model contains transfer costs, which are similar to turn penalties in the context of urban transportation networks (see Potts and Oliver, 1972). Spiess (1984) proposed a link labeling procedure (which is implemented in the EMME/2 code) for computing shortest paths on urban road networks with turn penalties at intersections. Unlike urban road networks, in our model a path that contains the links of more than one mode can be constructed only if a mode to mode tranfer is permitted at a node of the base network that belongs to the path.

In order to state the algorithm that we use, we introduce the following notation. Let c_a, $a \in A$ and c_t, $t \in T$ be the "lengths" of the arcs and transfers. In algorithm A, these correspond to a set of current marginal costs. The shortest path algorithm finds the shortest path from an origin $o \in O$ to all destinations $d \in D$ in the subnetworks defined by the subset of modes of $m(p) \in M(p)$. The shortest path is retraced by the means of pointers to the preceding link: b_d is the access link to destination d, b_a is the predecessor of link a; the length of the shortest paths are given by variables u_d and u_a, where u_a is the length of the shortest path from the origin o to link a inclusive of c_a; $\bar{A}$ represents the set of links which were examined (labeled) but do not yet have permanent labels.

Shortest path algorithm with transfers

Step 0 - Initialization

Lengths : $u_d = \infty$, $d \in D$, $u_o = 0$; $u_a = \infty$, $a \in A$

Predecessors : $b_d = -1$, $d \in D$; $b_o = 0$; $b_a = 0$, $a \in A$

Links to label : $\bar{A} = \emptyset$

Dummy arc : $\bar{a} = (\bar{i}, \bar{j}, \bar{m})$ with $\bar{i} = 0$, $\bar{j} = o$, $\bar{m} = m(p)$; $u_{\bar{a}} = 0$

Go to step 4.

Step 1 - Choice of arc to label

if $\bar{A} = \emptyset$ then STOP.

Choose $\bar{a} = (\bar{i}, \bar{j}, \bar{m})$ of A such that $u_{\bar{a}} < u_a$ for all $a \in \bar{A}$.

Arc $\bar{a}$ receives a permanent label: $\bar{A}=\bar{A}-\{\bar{a}\}$.

If $\bar{j}$ is a transfer node, go to step 3; if $\bar{j}$ is a regular node, go to step 4; otherwise continue.

Step 2 - Test of "head" node $\bar{j}$ (for destination node)

If $u_{\bar{a}} < u_{\bar{j}}$ then $u_{\bar{j}}=u_{\bar{a}}$; $b_{\bar{j}}=\bar{a}$.

Return to step 1.

Step 3 - Scan of successors with transfers

For each $a=(i,j,m)$ such that $i=\bar{j}$ and $m \epsilon m(p)$ do:

if there is a transfer $t=(\bar{a},a)$ do:

if $u_{\bar{a}} + c_t + c_a < u_a$ then $u_a=u_{\bar{a}} + c_t + c_a$; $b_a=\bar{a}$; $\bar{A}=\bar{A} \cup \{a\}$

otherwise, if $m=\bar{m}$ do:

if $u_{\bar{a}} + c_a < u_a$ then $u_a=u_{\bar{a}} + c_a$; $b_a=\bar{a}$; $\bar{A}=\bar{A} \cup \{a\}$

Return to step 1.

Step 4 - Scan of successors without transfers

For each $a=(i,j,m)$ such that $i=\bar{j}$ and $m=\bar{m}$ do:

if $u_{\bar{a}} + c_a < u_a$ then $u_a= u_{\bar{a}} + c_a$; $b_a=\bar{a}$, $\bar{A}=\bar{A} \cup \{a\}$

Return to step 1.

Several remarks concerning this shortest path algorithm are in order. Normally, each arc corresponds to only one mode. The dummy (connector) arc that is incident to the origin node o, has as its mode the set m(p) of permitted modes. This is done so that the first time that step 4 is carried out, all the arcs that leave o belong to the subnetwork. (This may be easily implemented by using an appropriate bit pattern to represent the permitted modes on a link and by using only one word in the computer implementation of the algorithm.) Thus, the first time that step 4 is executed, the test $m=\bar{m}$ is $m \epsilon \bar{m}=m(p)$. In step 3, the arcs are scanned in increasing order; therefore it is advantageous that the transfers of the type $(\bar{a},\cdot)$ are sorted in increasing order as well. The arcs that belong to the "forward star" of node i are sorted by mode in the arc table; thus an efficient test may preclude the scan of all the successor links in step 4. The

set $\bar{A}$ of labeled links is kept in the form of a binary heap, which makes it trivial to find the link $\bar{a}$ that is the next to be labelled.

We examine next the complexity of this shortest path algorithm. The following table shows the complexity of every step of the algorithm and the number of times that it is executed.

step	complexity	# times
0	$O(n_N + n_A)$	1
1	$O(\ln n_A)$	n_A
2	$O(1)$	$n_D \leq n_N$
3	$O(\Delta m(p))$	n_T
4	$O(\Delta m(p))$	$n_N - n_T$

The complexity of the entire algorithm for one origin node is then $O(n_A \ln n_A + \Delta m(p)\ n_N)$, where Δ is the average number of nodes connected to the current node by arcs leaving the node; since the first term dominates the second term in the parentheses, the complexity is then $O(n_A \ln n_A)$. As the shortest path algorithm must be applied for each origin and for each subset of modes $m(p) \in M(p)$. If n_p is the number of products, $|O|$ is the number of origin nodes and $|M(p)|$ is the number of mode subsets in a given problem then the complexity of the shortest path computations at each iteration of the solution algorithm proposed in section 4 is $O(n_p\ |M(p)|\,|O|\ n_A \ln n_A)$.

8. Computational Results

We first present in this section computational results obtained on a small network, which permits the comparison of the Gauss-Seidel - Linear Approximation algorithm (a) with the Linear Approximation Algorithm (b).

The test network consists of 5 origins, 5 destinations, 19 regular nodes, 116 links, and 226 transfers. Seven modes are permitted on the links of this network. The following test problems were constructed.

1: A single product is transported on the network and all modes are permitted. There is one origin/destination matrix. In this case (a) and (b) are equivalent as can be seen in Figure 5. The slight difference in the computational time is that in (a) the gap is recomputed at the end of a major cycle, whereas in (b) this gap is available automatically.

2: Three products are assigned to the network, where two products are permitted on all modes and the third is restricted to a subset of three modes. The demands of each of the three products are given by one matrix. In this case algorithm (a) is slightly more efficient than (b), as evidenced by Figure 6.

3: Five products are assigned to the network but only two may be transported by all the modes. The demand for the first product is given by three origin/destination matrices, each for different subsets of modes. Similarly, the demand for the second product is divided into two matrices. Figure 7 shows that the performance of algorithm (a) is very good.

4: The same five products of the previous test problem and the same origin/destination matrices are assigned to the network with the difference that all the modes are permitted. In addition, the generalized cost is the weighted sum of three costs (operation, delay, and energy), whereas in the previous problems, only the delay cost was considered. This has, as a consequence, the tripling of the evaluation time of the unit cost functions, which, for problems of this size, represent the largest part of the computational time. The results are shown in Figure 8.

We note that in all cases where there is more than one product algorithm (a) is more efficient than algorithm (b) both in terms of the number of iterations and the computational time (we note here that these results were obtained on an MC 68000 microcomputer not equipped with a floating point accelerator). Since algorithm (a) requires less central memory (RAM) than algorithm (b) it is well suited for problems of larger size.

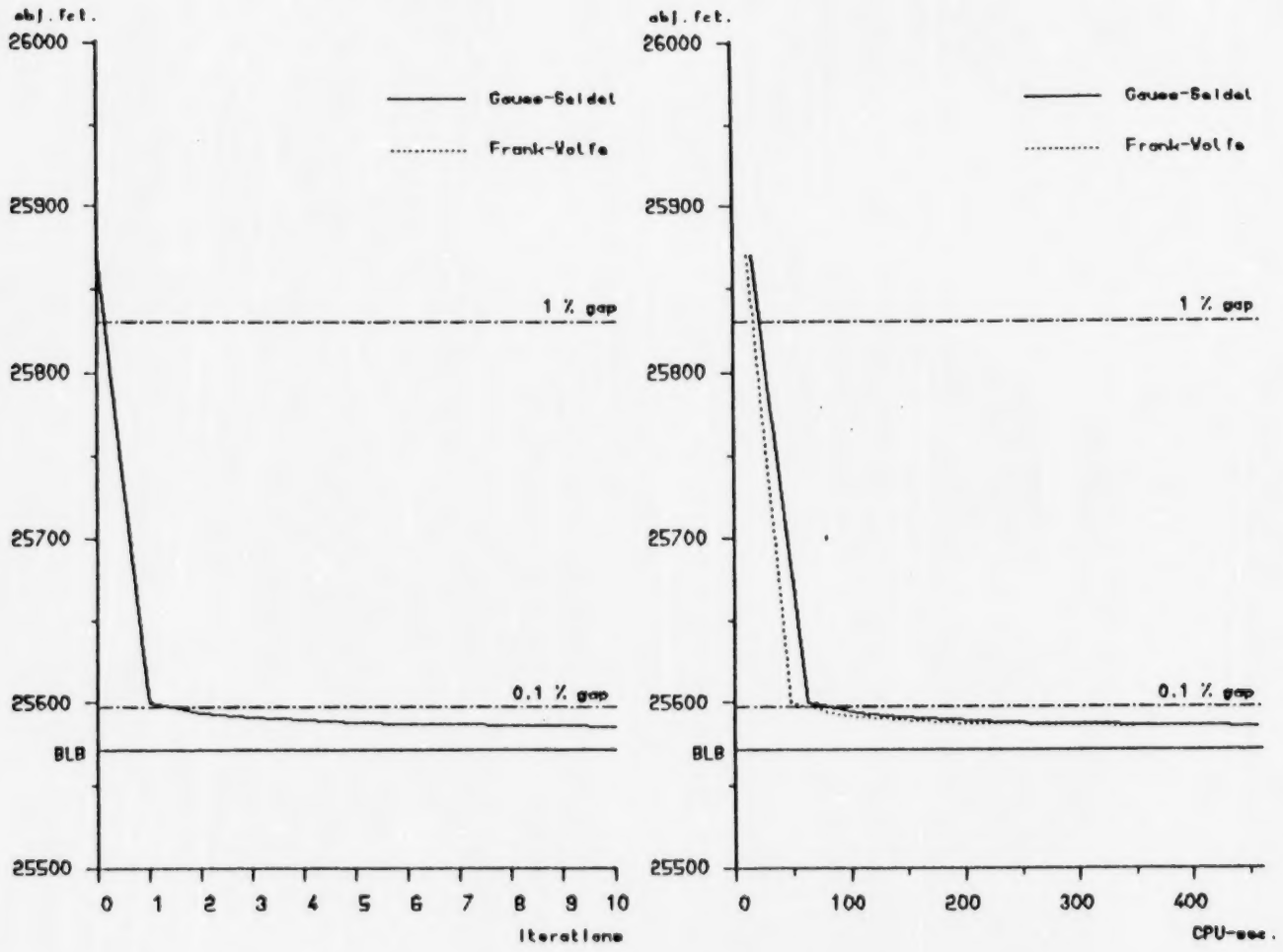


Figure 5: Problem 1

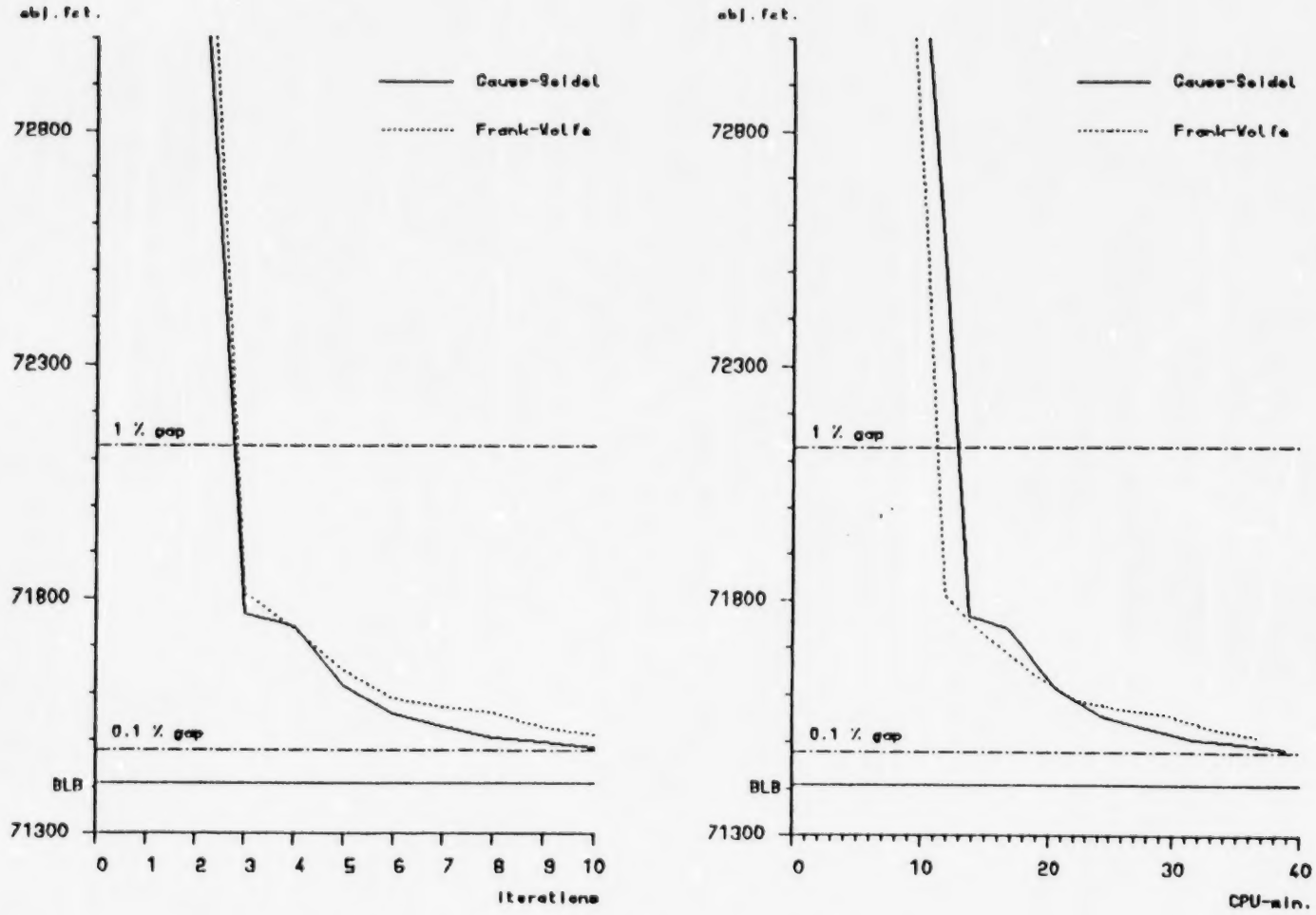


Figure 6: Problem 2

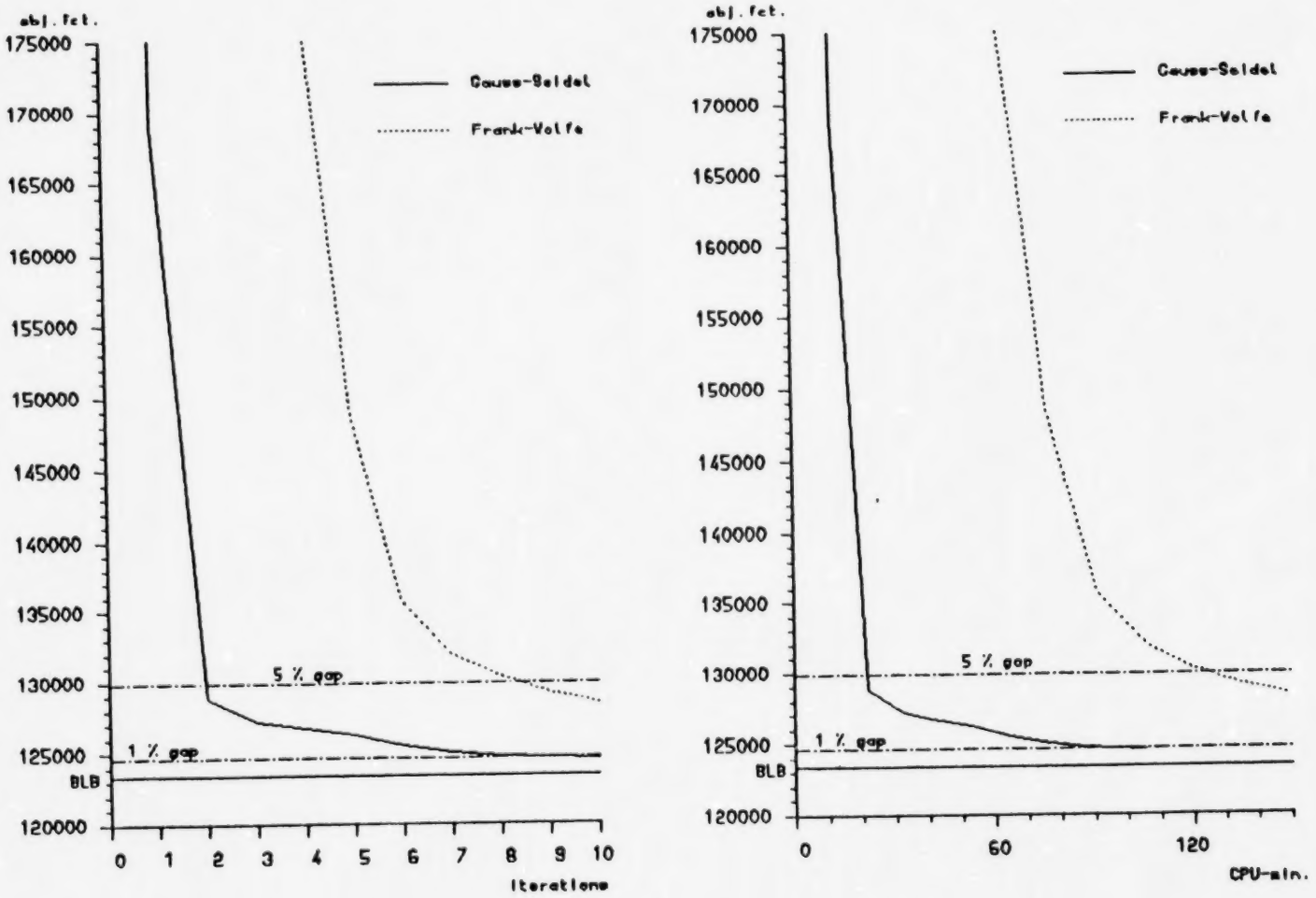
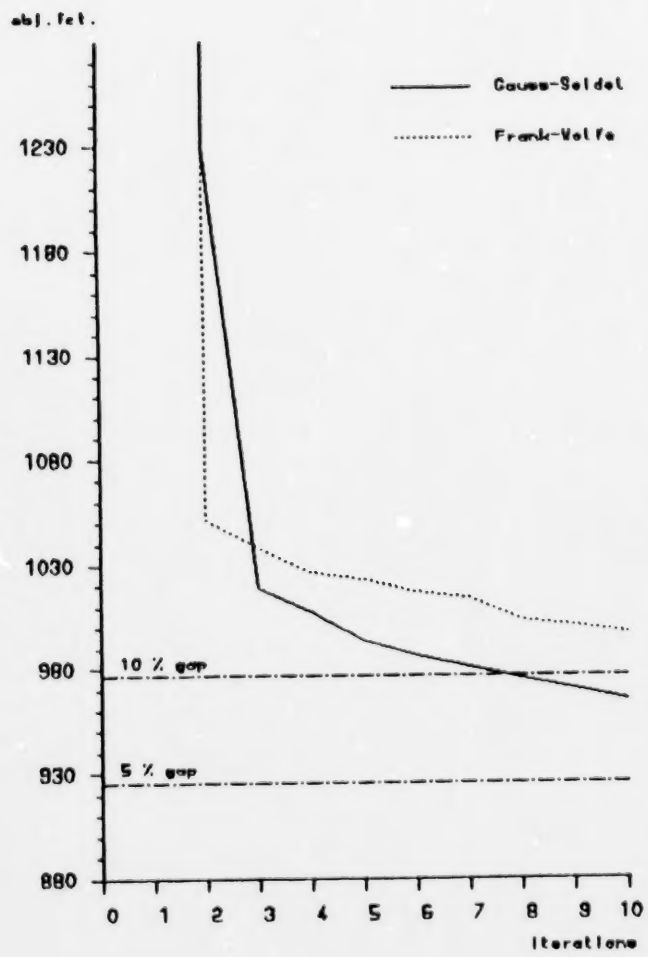


Figure 7: Problem 3



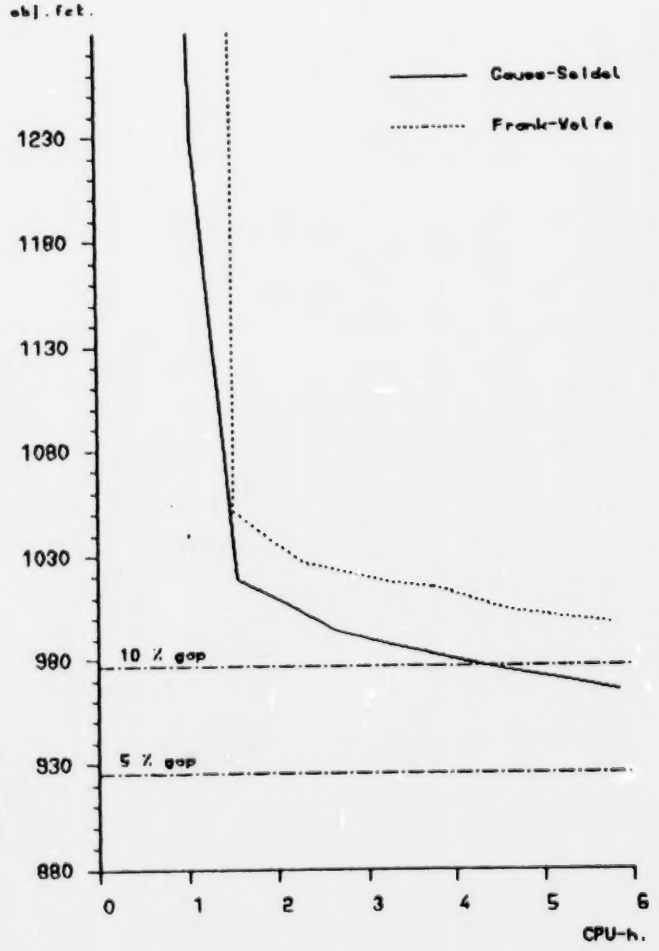


Figure 8: Problem 4

We conclude this section by giving the computational results obtained on the a' network that represents the national freight transportation network of Brazil. This network consists of 211 origins and destinations, 1243 regular nodes, 4957 links and, 5718 transfers. Six products (cement and steel products, iron ore, fertilizer and coal, soya oil, soya grain, and all others) are assigned and 10 modes are permitted on the links of this network. Only the delay cost was considered. The test was carried out on a SUN/3 work station (MC 68020 equipped with DC 68881 floating point accelerator operating at 16.7 Mhz).

Figures 9 and 10 show the decrease of the objective function vs the number of iterations and the total computation time respectively. Figures 11 and 12 show the decrease of the relative gap vs the number of interations and the total computation time respectively.

While the computational time per iteration appears to be high, approximately 80% of the time of an iteration is spent on computing the marginal costs and the gap. This is due to our choice of making the code "user friendly" and require the user to enter the cost functions directly and not impose on time the calculation of the analytical form of the cost functions.

9. Conclusions

We have developed a multimode multiproduct network assignment model for strategic planning of freight flows, that permits the modeling of national and regional freight transportation networks. The multiproduct aspects of the model are exploited in the solution procedure, which is a Gauss-Seidel Linear Approximation Method. The computational results obtained demonstrate that the method is applicable to large multimodal transportation networks.

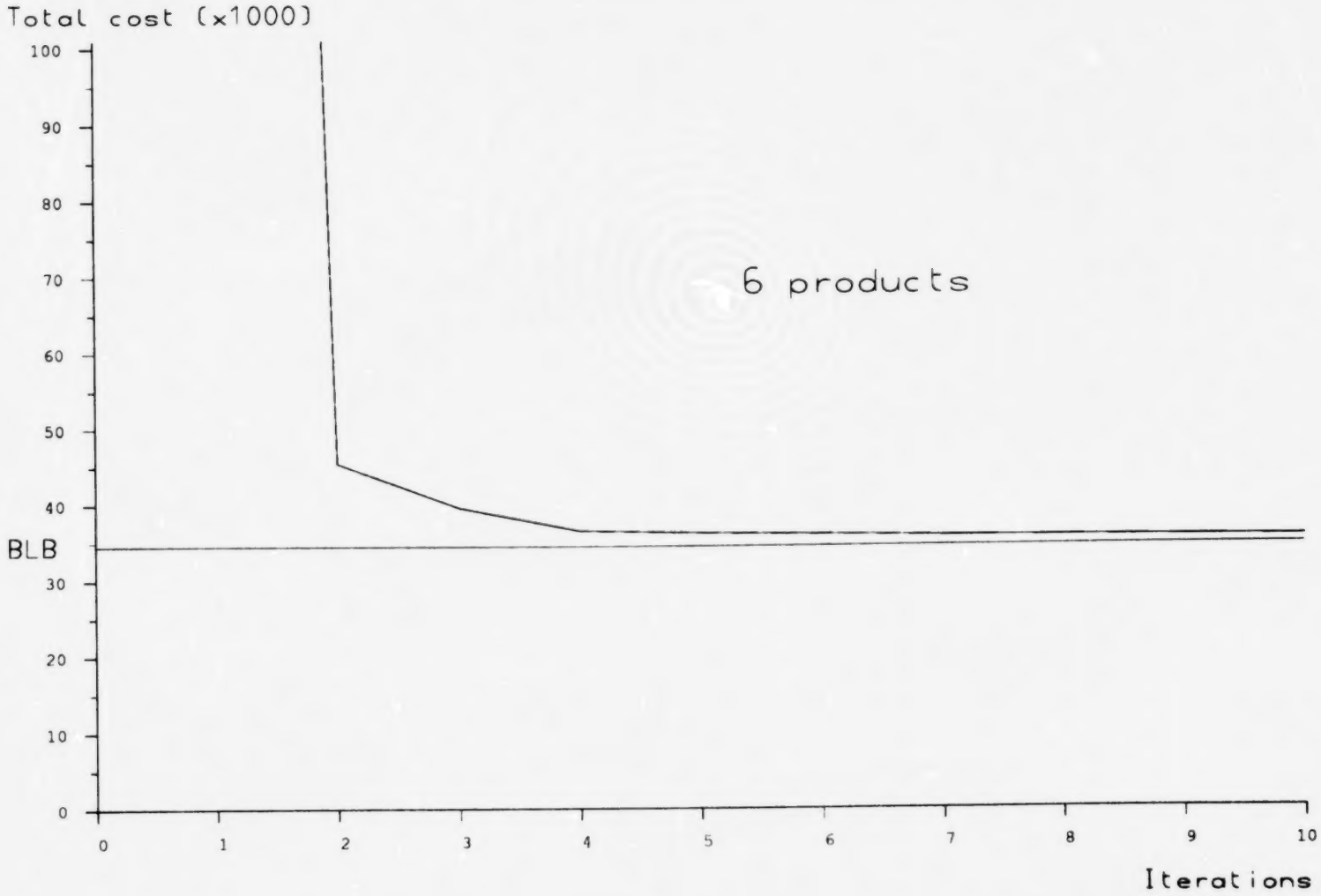

Total cost (x1000)
100
90
80
70
60
50
40
BLB
30
20
10
0
6 products
0
1
2
3
4
5
6
7
8
9
10
Iterations

Figure 9

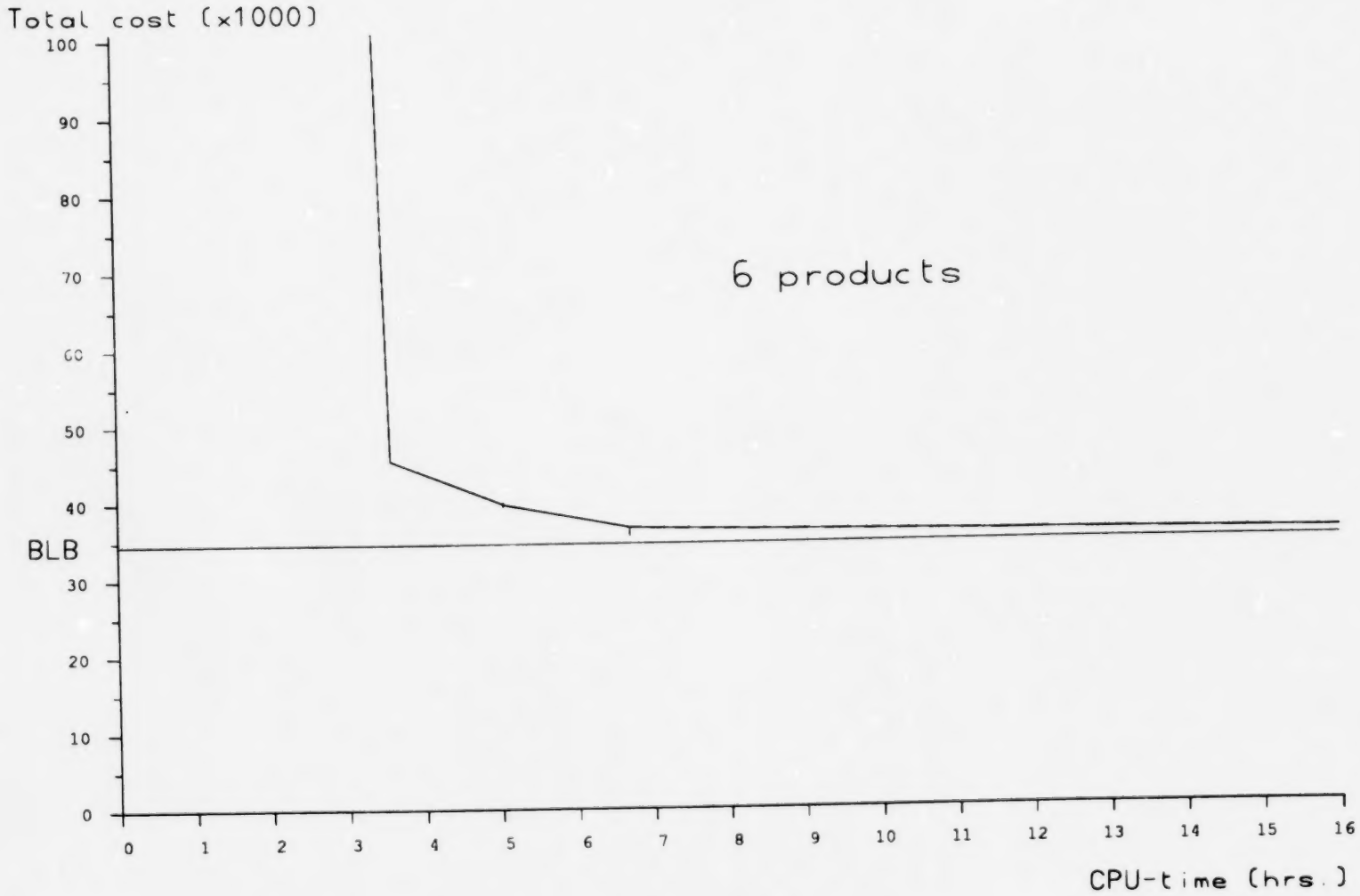

Total cost (x1000)
100
90
80
70
60
50
40
BLB
30
20
10
0
6 products
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
CPU-time (hrs.)

Figure 10

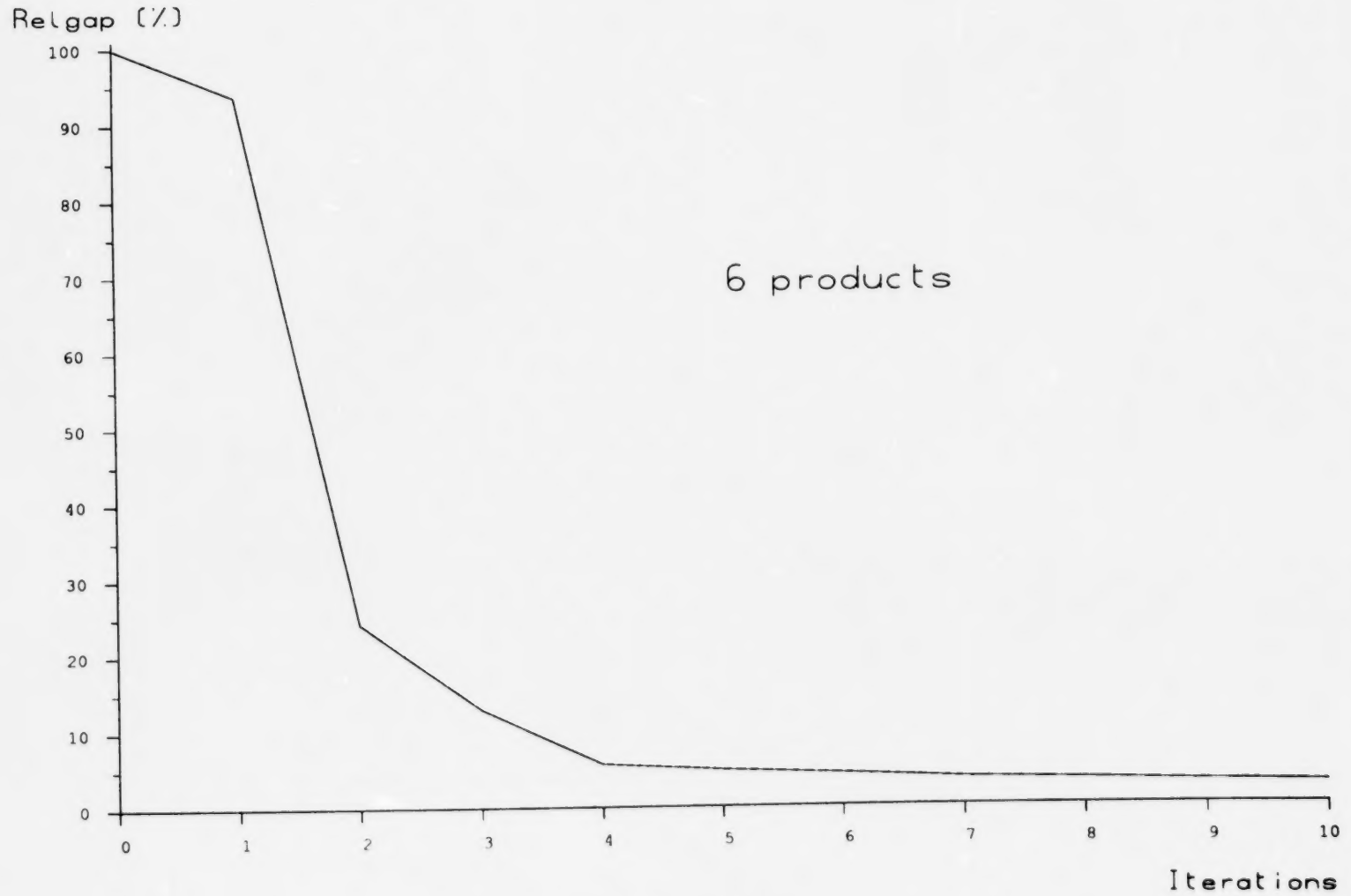

Relgap (%)
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
Iterations
6 products

Figure 11

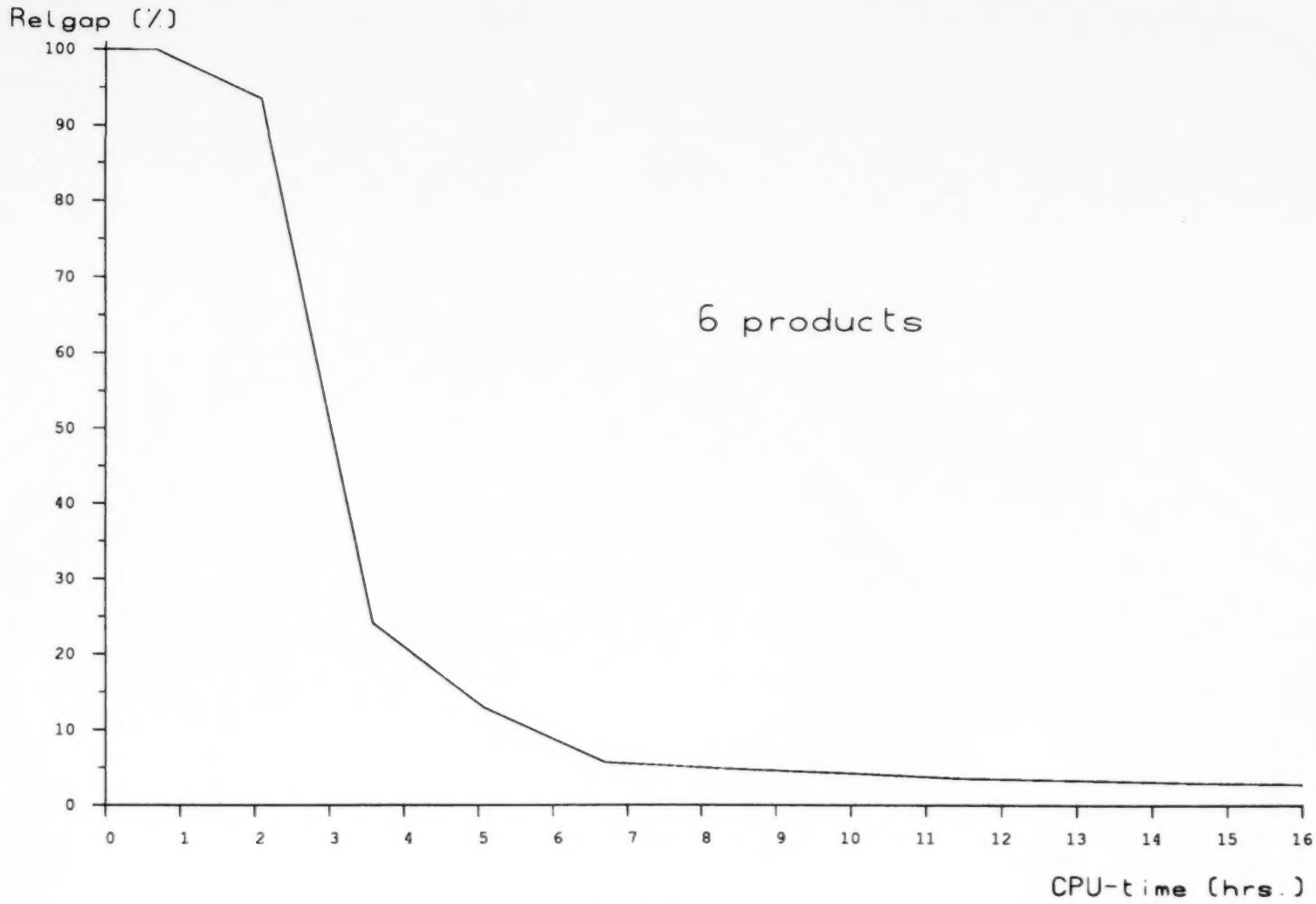

Relgap (%)
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
6 products
CPU-time (hrs.)

Figure 12

REFERENCES

BRONZINI, M.S. (1980) "Evolution of a Multimodal Freight Transportation Network Model", Proceedings of the 21st Annual Meeting Transportation Research Forum, 475-485.

CASALETTO, J.J., RICE, J.R. (1978) "Smoothing and Estimating Derivatives of Equispaced Data", Applicable Analysis 8:143-152.

CRAINIC, T.G., FLORIAN, M., LEAL E. (1987) "A Model for the Strategic Planning of Freight Transportation by Rail", publication #518, Centre de recherche sur les transports, Université de Montréal.

FLORIAN, M., LOS, M. (1982) "A New Look at Static Price Equilibrium Models", Regional Science and Urban Economics 12:579-597.

FLORIAN, M. (1986) "Nonlinear Cost Network Models in Transportation Analysis", Mathematical Programming Study 26:167-196.

FRANK, M., WOLFE, P. (1956) "An Algorithm for Quadratic Programming", Naval Research Logistics Quarterly 3:95-110.

FRIESZ, T.L., TOBIN, R.L., HARKER, P.T. (1983) "Predictive Intercity Freight Network Models", Transportation Research 17A:409-417.

FRIESZ, T.L., GOTTFRIED, J.A., MORLOK, E.K. (1986) "A Sequential Shipper-Carrier Network Model for Predicting Freight Flows", Transportation Science 20:80-91.

GUÉLAT, J. (1986a) "Calcul des dérivées (partielles) des fonctions de coût dans le progiciel STAN", publication #535, Centre de recherche sur les transports, Université de Montréal.

GUÉLAT, J. (1986b) "Une procédure de résolution du problème d'affectation multimode-multiproduct par décomposition sur un espace produit", publication #488, Centre de recherche sur les transports, Université de Montréal.

JONES, P.S., SHARP, G.P. (1979) "Multi-Mode Intercity Freight Transportation Planning for Underdeveloped Regions", Proceedings of the 20th Annual Meeting Transportation Research Forum.

KELLISON, S.G. (1975) "Fundamentals of Numerical Analysis", Richard D. Irwin Inc., Homewood, Illinois.

KRESGE, D.T., ROBERTS, P.O. (1971) "System Analysis and Simulation Models", in Techniques of Transport Planning, Meyers Jr. (ed), vol. 2, Brookings Institute, Washington, D.C.

McGINNIS, L.F., SHARP, G.P., YU DHC (1981) "Procedures for Multi-State, Multi-Mode Analysis: Vol. IV", Transportation Modeling and Analysis, U.S., D.O.T. report no. DOT-OST-80050-17/V.N.

POTTS, R.B., OLIVER, R.M. (1972) "Flows in Transportation Networks", Academic Press, NY.

SAMUELSON, P.A. (1952) "Spatial Price Equilibrium and Linear Programming", American Economic Review 42:283-303.

SHARP, G.P. (1979) "A Multi-Commodity Intermodal Transportation Model", Proceedings of the 20th Annual Meeting Transportation Research Forum.

SPIESS, H. (1984) "Contributions à la théorie et aux outils de planification des réseaux de transports urbains", Ph.D. thesis, Département d'Informatique et de Recherche opérationnelle, publication #382, Centre de recherche sur les transports, Université de Montréal.

TAKAYAMA, T., JUDGE, G.G. (1964) "Equilibrium among Spatially Separated Markets: A Reformulation", Econometrica 32:510-524.

TAKAYAMA, T., JUDGE, G.G. (1970) "Alternative Spatial Price Equilibrium Models", Journal of Regional Science 10:1-12.

ZANGWILL, W.I. (1969) "Nonlinear Programming - A Unified Approach", Prentice Hall, Englewood Cliffs, NJ.